

# Geometric ergodicity of Rao and Teh's algorithm for Markov jump processes

Błażej Miasojedow<sup>a</sup>, Wojciech Niemiro<sup>a</sup>

<sup>a</sup>*Institute of Applied Mathematics, University of Warsaw  
Banacha 2, 02-097 Warsaw, Poland*

---

## Abstract

Rao and Teh (2013) introduced an efficient MCMC algorithm for sampling from the posterior distribution of a hidden Markov jump process. The algorithm is based on the idea of sampling virtual jumps. In the present paper we show that the Markov chain generated by Rao and Teh's algorithm is geometrically ergodic. To this end we establish a geometric drift condition towards a small set.

*Keywords:* Continuous time Markov processes, MCMC, Hidden Markov models, Posterior sampling, Geometric ergodicity, Drift condition, Small set

---

## 1. Introduction

Markov jump processes (MJP) are natural extension of Markov chains to continuous time. They are widely applied in modelling of the phenomena of chemical, biological, economic and other sciences.

In many applications it is necessary to consider a situation where the trajectory of a MJP is not observed directly, only partial and noisy observations are available. Typically, the posterior distribution over trajectories is then analytically intractable. In the literature there exist several approaches to the above mentioned problem: based on sampling (Boys et al., 2008; El-Hay et al., 2008; Fan and Shelton, 2008; Fearnhead and Sherlock, 2006; Golightly and Wilkinson, 2011; Golightly and Wilkinson, 2014; Golightly et al., 2015; Nodelman et al., 2002; Rao and Teh, 2013, 2012), and also based on

---

*Email addresses:* `bmia@mimuw.edu.pl` (Błażej Miasojedow), `wniem@mimuw.edu.pl` (Wojciech Niemiro)

numerical approximations. To the best of our knowledge the most general, efficient and exact method for a finite state space is that proposed by [Rao and Teh \(2013\)](#), and extended to a more general class of continuous time discrete systems in [Rao and Teh \(2012\)](#).

In the present paper we establish geometric ergodicity of Rao and Teh’s algorithm for homogeneous MJPs. Geometric ergodicity is a key property of Markov chains which allows to use Central Limit Theorem for sample averages.

The rest of the paper is organised as follows. In Section 2 we briefly introduce hidden Markov jump processes, next in Section 3 we recall the Rao and Teh’s algorithm. The main result is proved in Section 4.

## 2. Hidden Markov jump processes

Consider a continuous time homogeneous Markov process  $\{X(t), t^{\min} \leq t \leq t^{\max}\}$  on a finite state space  $\mathcal{S}$ . Its probability law is defined via the initial distribution  $\nu(s) = \mathbb{P}(X(t^{\min}) = s)$  and the transition intensities

$$Q(s, s') = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}(X(t+h) = s' | X(t) = s)$$

for  $s, s' \in \mathcal{S}$ ,  $s \neq s'$ . Let  $Q(s) = \sum_{s' \neq s} Q(s, s')$  denote the intensity of leaving state  $s$ . For definiteness, assume that  $X$  has right continuous trajectories. We say  $X$  is a Markov jump process (MJP).

Suppose that process  $X$  cannot be directly observed but we can observe some random quantity  $Y$  with probability distribution  $L(Y|X)$ . Let us say  $Y$  is the evidence and  $L$  is the likelihood. The problem is to restore the hidden trajectory of  $X$  given  $Y$ . From the Bayesian perspective, the goal is to compute/approximate/sample from the posterior

$$p(X|Y) \propto p(X)L(Y|X).$$

Function  $L$ , transition probabilities  $Q$  and initial distribution  $\nu$  are assumed to be known. To get explicit form of posterior distribution we consider a typical form of noisy observation. Assume that the trajectory  $X([t^{\min}, t^{\max}])$  is observed independently at  $k$  deterministic time points with some random errors. Formally, we observe  $Y = (Y_1, \dots, Y_k)$  where

$$L(Y|X) = \prod_{j=1}^k L_j(Y_j|X(t_j^{\text{obs}})), \quad (1)$$

for some fixed known points  $t^{\min} \leq t_1^{\text{obs}} < \dots < t_r^{\text{obs}} \leq t^{\max}$ .

### 3. Uniformization and Rao and Teh's algorithm

The so-called “uniformization” is a classical and well-known representation of a Markov jump process in terms of potential times of jumps and the corresponding states (Jensen, 1953). Every trajectory  $X([t^{\min}, t^{\max}])$  is right continuous and piece-wise constant:  $X(t) = S_{i-1}$  for  $T_{i-1} \leq t < T_i$ , where random variables  $T_i$  are such that  $t^{\min} < T_1 < \dots < T_N < t^{\max}$  (by convention,  $T_0 = t^{\min}$  and  $t^{\max} < T_{N+1}$ ). Random sequence of states  $S = (S_1, \dots, S_N)$  such that  $S_i = X(T_i)$  is called a skeleton. We do not assume that  $S_{i-1} \neq S_i$ , and therefore the two sequences

$$\begin{pmatrix} T \\ S \end{pmatrix} = \begin{pmatrix} t^{\min} & T_1 & \dots & T_i & \dots & T_N & t^{\max} \\ S_0 & S_1 & \dots & S_i & \dots & S_N \end{pmatrix}.$$

represent the process  $X$  in a redundant way: many pairs  $(T, S)$  correspond to the same trajectory  $X([t^{\min}, t^{\max}])$ . Let  $J = \{i \in [1 : N] : S_{i-1} \neq S_i\}$ , so that  $T_J = (T_i : i \in J)$  are moments of *true* jumps and  $T_{-J} = T \setminus T_J = (T_i : i \notin J)$  are *virtual* jumps. We write  $[l : r] = \{l, l+1, \dots, r\}$ . By a harmless abuse of notation, we identify increasing sequences of points in  $[t^{\min}, t^{\max}]$  with finite sets. Note that the trajectory of  $X$  is uniquely defined by  $(T_J, S_J)$ . Let us write  $X \equiv (T_J, S_J)$  and also use the notation  $J(X) = T_J$  for the set of true jumps.

Uniformization obtains if  $T$  is a (sequence of consecutive points of) a homogeneous Poisson process with intensity  $\lambda$ , where  $\lambda \geq Q^{\max} = \max_s Q(s)$ . The skeleton  $S$  is then (independently of  $T$ ) a discrete time homogeneous Markov chain with the initial distribution  $\nu$  and the transition matrix

$$P(s, s') = \begin{cases} \frac{Q(s, s')}{\lambda} & \text{if } s \neq s' ; \\ 1 - \frac{Q(s)}{\lambda} & \text{if } s = s' . \end{cases} \quad (2)$$

Rao and Teh (2013) exploit uniformization to construct a special version of Gibbs sampler which converges to the posterior  $p(X|Y)$ . The key facts behind their algorithm are the following. First, given the trajectory  $X \equiv (T_J, S_J)$  the conditional distribution of virtual jumps  $T_{-J}$  is that of the non-homogeneous (actually piece-wise homogeneous) Poisson process with intensity  $\lambda - Q(X(t)) \geq 0$ . Second, this distribution does not change if we introduce the likelihood. Indeed,  $L(Y|X) = L(Y|T_J, S_J)$ , so  $Y$  and  $T_{-J}$

are conditionally independent given  $(T_J, S_J)$  and thus  $p(T_{-J}|T_J, S_J, Y) = p(T_{-J}|T_J, S_J)$ . Third, the conditional distribution  $p(S|T, Y)$  is that of a hidden discrete time Markov chain and can be efficiently sampled from using the algorithm FFBS (Forward Filtering-Backward Sampling).

The Rao and Teh's algorithm generates a Markov chain  $X_0, X_1, \dots, X_m, \dots$  (where  $X_m = X_m([t^{\min}, t^{\max}])$  is a trajectory of MJP), convergent to  $p(X|Y)$ . A single step, that is the rule of transition from  $X_{m-1} = X$  to  $X_m = X'$  is described in Algorithm 1.

---

**Algorithm 1** Single step of Rao and Teh's algorithm.

---

**input:** previous state  $(T_J, S_J) \equiv X$  and observation  $Y$ .  
(V) Sample a Poisson process  $V$  with intensity  $\lambda - Q(X(t))$  on  $[t^{\min}, t^{\max}]$ .  
Let  $T' = T_J \cup V$  *{new set of potential times of jumps}*.  
(S) Draw new skeleton  $S'$  from the conditional distribution  $p(S'|T', Y)$  by FFBS. The new allocation of virtual and true jumps is via  $J' = \{i : S'_{i-1} \neq S'_i\}$  *{we discard new virtual jumps  $T'_{-J'}$ }*.  
**return** new state  $(T'_{J'}, S'_{J'}) \equiv X'$ .

---

Convergence of the algorithm has been shown by its authors in [Rao and Teh \(2013\)](#). It follows from the fact that the chain has the stationary distribution  $p(X|Y)$  and is irreducible and aperiodic, provided that  $\lambda > Q^{\max}$ .

#### 4. Main result

Consider the Markov chain  $X_0, X_1, \dots, X_m, \dots$  generated by the Rao and Teh's algorithm. Its transition kernel is denoted by  $A$ . Put differently,  $A(X, dX')$  is the probability distribution corresponding to Algorithm 1. Let  $\Pi(dX)$  be the posterior distribution of  $X$  given  $Y$ . In this paper we consider only Monte Carlo randomness, so  $Y$  is fixed and can be omitted in notation. The obvious standing assumption is that  $L(Y|X) > 0$  happens with nonzero probability if  $X$  is given by  $\nu$  and  $Q$ . It just means that the hidden MJP under consideration is "possible".

**Theorem 1.** *Assume that the matrix of intensities  $Q$  is irreducible and  $\lambda > Q^{\max}$ . Then the chain is geometrically ergodic. There exist constant  $\gamma < 1$  and function  $M$  such that for every  $X$ ,*

$$\|A^m(X, \cdot) - \Pi(\cdot)\|_{\text{tv}} \leq \gamma q^m M(X).$$

We begin with some auxiliary results.

**Lemma 1.1.** *Let  $S_0, S_1, \dots, S_n$  be a Markov chain on a finite state space  $\mathcal{S}$ , with transition matrix  $P$ . Assume that  $P$  is irreducible and  $P(s, s) \geq \eta > 0$  for all  $s \in \mathcal{S}$ . Let  $J = \{i \in [1: n] : S_i \neq S_{i-1}\}$ . There exist  $n_0$  and  $\delta > 0$  such that for every  $n \geq n_0$  and for all  $s_0, s_n \in \mathcal{S}$ ,*

$$\mathbb{E}(|J| | S_0 = s_0, S_n = s_n) \leq (1 - \delta)n.$$

PROOF. Under the assumptions of the lemma, the Markov chain is ergodic. Let  $\pi$  be its (strictly positive) stationary distribution. Fix  $\varepsilon > 0$ . There exists  $n_0$  such that for  $n \geq n_0/2 - 1$  we have for all  $s_0, s_n \in \mathcal{S}$ ,

$$(1 - \varepsilon)\pi(s_n) \leq \mathbb{P}(S_n = s_n | S_0 = s_0) \leq (1 + \varepsilon)\pi(s_n).$$

Now if  $n \geq n_0$  then we can choose  $n_1$  such that  $n_1 \geq n_0/2$  and  $n - n_1 \geq n_0/2 - 1$ . Let  $J_1 = \{i \in [1: n_1] : S_i \neq S_{i-1}\}$ . With the notation  $\rho(j, s) = \mathbb{P}(|J_{n_1}| = j, S_{n_1} = s | S_0 = s_0)$  we have

$$\begin{aligned} \mathbb{E}(|J_1| \mathbb{I}(S_n = s_n) | S_0 = s_0) &= \sum_j \sum_s j \rho(j, s) P^{n-n_1}(s, s_n) \\ &\leq \sum_j \sum_s j \rho(j, s) (1 + \varepsilon) \pi(s_n) \\ &= \mathbb{E}(|J_{n_1}| | S_0 = s_0) (1 + \varepsilon) \pi(s_n) \\ &\leq (1 - \eta) n_1 (1 + \varepsilon) \pi(s_n). \end{aligned}$$

The last inequality follows from the fact that  $\mathbb{P}(S_i = s | S_{i-1} = s) \geq \eta$ . Consequently,

$$\begin{aligned} \mathbb{E}(|J| | S_n = s_n, S_0 = s_0) &= \frac{\mathbb{E}(|J_1| \mathbb{I}(S_n = s_n) | S_0 = s_0)}{\mathbb{P}(S_n = s_n | S_0 = s_0)} \\ &\leq (1 - \eta) \frac{1 + \varepsilon}{1 - \varepsilon} n_1 = (1 - 2\delta) n_1, \end{aligned}$$

where  $\delta > 0$  if  $\varepsilon$  is chosen sufficiently small. Finally, since  $|J| \leq |J_1| + n - n_1$  and  $n_1 \geq n/2$ , we obtain

$$\mathbb{E}(|J| | S_n = s_n, S_0 = s_0) \leq (1 - 2\delta) n_1 + n - n_1 = n - 2\delta n_1 \leq (1 - \delta)n.$$

We get the conclusion.

**Lemma 1.2.** *Let the assumptions of Lemma 1.1 hold. Numbers  $n_0$  and  $\delta$  are the same as in this lemma. Let  $k$  be fixed. If  $n \geq (k+1)n_0$ , then for arbitrarily chosen indices  $0 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq n$  and states  $s_1, \dots, s_k$  we have that*

$$\mathbb{E}(|J| | S_{i_1} = s_1, \dots, S_{i_k} = s_k) \leq \left(1 - \frac{\delta}{k+1}\right) n,$$

*provided that  $\mathbb{P}(S_{i_1} = s_1, \dots, S_{i_k} = s_k) > 0$ .*

PROOF. Write  $n_j = i_j - i_{j-1}$ , where by convention  $i_0 = 0$  and  $i_{k+1} = n$ . Let  $J^{(j)} = \{i \in [i_{j-1}+1: i_j] : S_i \neq S_{i-1}\}$  for  $j = 1, \dots, k+1$ . (Let us mention that we have not excluded the case when  $i_{j-1} = i_j$  for some  $j$  and thus  $J^{(j)} = \emptyset$ . Of course, then we must have  $s_{j-1} = s_j$  in the conditional expectation.) If  $n \geq (k+1)n_0$  then there is at least one  $j$  such that  $n_j \geq n/(k+1) \geq n_0$ , because  $\sum_j n_j = n$ . From Lemma 1.1 we infer that for this  $j$  we have

$$\mathbb{E}(|J^{(j)}| | S_{i_1} = s_1, \dots, S_{i_k} = s_k) = \mathbb{E}(|J^{(j)}| | S_{i_{j-1}} = s_{j-1}, S_{i_j} = s_j) \leq (1 - \delta)n_j.$$

The equality above is true because of the (two-sided) Markov property. Analogously as in the previous proof, we observe that  $|J| \leq |J^{(j)}| + n - n_j$ , so

$$\mathbb{E}(|J| | S_{i_1} = s_1, \dots, S_{i_k} = s_k) \leq (1 - \delta)n_j + n - n_j = n - \delta n_j \leq \left(1 - \frac{\delta}{k+1}\right) n.$$

In the next proposition we establish a geometric drift condition for the Markov chain  $X_0, X_1, \dots, X_m, \dots$ . Consider a single step, that is transition from  $X_{m-1} = X$  to  $X_m = X'$ . Thus only Monte Carlo randomness in Algorithm 1 is taken into account. The dependence on the input trajectory  $X$  (and on  $Y$ ) is implicitly assumed but indicated only when necessary. Recall that  $|J(X)|$  is the number of true jumps of the trajectory  $X([t^{\min}, t^{\max}])$ .

**Proposition 1.1.** *There exist  $q < 1$  and  $c < \infty$  such that in a single step of the Rao and Teh's algorithm we have  $\mathbb{E}(|J(X')| | X) \leq q|J(X)| + c$ .*

PROOF. Let us analyse what happens in both two stages of Algorithm 1. We fix the initial  $X$ . In the first stage we add a new set  $V$  of potential jumps. Since  $|V|$  has the Poisson distribution with intensity  $\int_{t^{\min}}^{t^{\max}} (\lambda - Q(X(t))) dt$ , we have  $\mathbb{E}|V| \leq \mu = \lambda(t^{\max} - t^{\min})$ . Thus we obtain  $T'$  with  $\mathbb{E}|T'| \leq |J(X)| + \mu$ . In the second stage  $T'$  is “thinned” to  $T'_j$ . We will prove that

$$\mathbb{E}(|J'| | T') \leq \left(1 - \frac{\delta}{k+1}\right) |T'|, \text{ if } |T'| \geq (k+1)n_0. \quad (3)$$

Inequality (3) implies  $\mathbb{E}(|J(X')|) = \mathbb{E}(|J'|) \leq (1 - \delta/(k+1))(|J(X)| + \mu)$ . The conclusion of the theorem follows with  $q = (1 - \delta/(k+1))$ . To ensure that the conclusion holds also if  $|T'| < (k+1)n_0$ , we can choose  $c = q\mu + (k+1)n_0$  so that  $|J'| < c$ .

It remains to show (3). We consider the second stage of Algorithm 1 (sampling a new skeleton  $S'$  from  $p(S'|T', Y)$ ). From now on, the result of the first stage (updating  $T$  to  $T'$ ) is fixed. Note that our assumption about the structure of observations (1) implies that  $p(S'|T', Y) \propto p(S'|T') \prod_{j=1}^r L_j(Y_j|S'_{i_j})$ , where  $i_j = \max\{i : T'_i \leq t_j^{\text{obs}}\}$ . Since  $p(S'|T')$  is the distribution of a Markov chain, it follows that

$$p(S'|T', Y) \propto \nu(S'_0) \prod_{i=1}^{|T'|} P(S'_{i-1}, S'_i) \prod_{j=1}^k L_j(S'_{i_j}),$$

where  $P$  is given by (2) and  $L_j(s) = L_j(Y_j|s)$ .

Although the actual sampling from  $p(S'|T', Y)$  is by FFBS, exactly the same result can be obtained via rejection sampling as follows.

- (S1) Simulate Markov chain  $S'$  (of length  $|T'|$ ) with transition matrix  $P$  given by (2) and initial distribution  $\nu$ .
- (S2) The skeleton  $S'$  is accepted with probability  $\prod_{j=1}^k (L_j(S'_{i_j})/L_j^{\max})$ , with some  $L_j^{\max} \geq \max_s L_j(s)$ . Otherwise go to (S1).

(Of course the rejection method is highly inefficient and is considered only to clarify presentation.) Now we are in a position to use Lemma 1.2, with  $\eta = \lambda - Q^{\max}$ . If  $|T'| \geq (k+1)n_0$  then

$$\begin{aligned} \mathbb{E}(|J'| | T') &= \mathbb{E}(|J'| | S' \text{ accepted}) \\ &= \sum_{s_1, \dots, s_k} \mathbb{E}(|J'| | S'_{i_1} = s_1, \dots, S'_{i_k} = s_k) \mathbb{P}(S'_{i_1} = s_1, \dots, S'_{i_k} = s_k | S' \text{ accepted}) \\ &\leq \left(1 - \frac{\delta}{k+1}\right) |T'| \sum_{s_1, \dots, s_k} \mathbb{P}(S'_{i_1} = s_1, \dots, S'_{i_k} = s_k | S' \text{ accepted}) \\ &\leq \left(1 - \frac{\delta}{k+1}\right) |T'| \end{aligned}$$

(of course in the sum above we can omit “impossible sequences”  $s_1, \dots, s_k$ ). We have proved (3) and we are done.

**Remark 1.1.** The proof of Proposition 1.1 could be significantly simpler if we assumed that the likelihood  $L_j(s)$  is strictly positive. However, this assumption is not satisfied in many interesting applications (e.g. if the observations are without noise).

**Proposition 1.2.** *The set  $\{X : |J(X)| \leq h\}$  is 1-small for every  $h$ .*

PROOF. We are going to construct a skeleton  $s^*$  and a set of the corresponding times of jumps  $\mathcal{T}$  which are achievable with uniformly lower bounded probability, in one step of the algorithm, from any trajectory  $X$  such that  $|J(X)| \leq h$ . Let us first choose a sequence  $s^\dagger = (s_1^\dagger, s_2^\dagger, \dots, s_k^\dagger)$  such that

$$\prod_{j=1}^k (L_j(s_j^\dagger) / L_j^{\max}) = \beta_1 > 0 .$$

By irreducibility of kernel  $P$  we can extend  $s^\dagger$  to a possible skeleton  $s^*$ , i.e. we define a sequence  $s^* = (s_0^*, \dots, s_l^*)$  for some  $l \geq k$  such that  $s^\dagger$  is a subsequence of  $s^*$ ,  $s_{i-1}^* \neq s_i^*$  and

$$\nu(s_0^*) \prod_{i=1}^l P(s_i^*, s_{i+1}^*) = \beta_2 > 0 .$$

We denote by  $i_1, \dots, i_k$  indices of elements of  $s^\dagger$  in  $s^*$ . Now we define a set  $\mathcal{T}$  as follows.

$$\mathcal{T} = \{(t_1, \dots, t_l) : t_{i_j} \leq t_j^{\text{obs}}, t_{i_{j+1}} > t_j^{\text{obs}} \text{ for } j = 1, \dots, k\} .$$

Fix  $X$  with  $|J(X)| \leq h$ . Let us first describe a special way in which Algorithm 1 can be executed. In stage (V) we can independently sample two Poisson processes on the interval  $[t^{\min}, t^{\max}]$ , say  $V^0$  and  $V^{\text{rest}}$ , with intensities  $\lambda - Q^{\max}$  and  $Q^{\max} - Q(X(t))$ , respectively. Next set  $V = V^0 \cup V^{\text{rest}}$  and  $T' = J(X) \cup V$ . Note that

$$\mathbb{P}(V^0 \in \mathcal{T}) = \beta_3 > 0 .$$

In stage (S) of the Algorithm 1 we construct skeleton  $S'$ . Assume that we use rejection sampling, defined by (S1) and (S2) in the previous proof. We use the notation as in Algorithm 1. Recall that  $J' = \{i : S'_{i-1} \neq S'_i\}$ . We will bound from below the probability that at stage (S1) all points belonging



to  $V^{\text{rest}} \cup J(X)$  are changed to virtual jumps, while jumps at  $V^0$  form the skeleton  $s^*$ . We have

$$\begin{aligned}\mathbb{P}(T'_{J'} = V^0, S'_{J'} = s^* | T') &\geq \beta_2 \sum_{k=0}^{\infty} \eta^{|J(X)|+k} \mathbb{P}(V^{\text{rest}} = k | X) \\ &= \beta_2 \mathbb{E}(\eta^{h+|V^{\text{rest}}|} | X) \geq \beta_2 \eta^{h+\mathbb{E}(|V^{\text{rest}}| | X)}\end{aligned}$$

by Jensen inequality. Since  $V^{\text{rest}}$  is a Poisson process with intensity bounded by  $Q^{\max}$  we have  $\mathbb{E}(|V^{\text{rest}}| | X) \leq Q^{\max}(t^{\max} - t^{\min})$ , therefore we conclude that  $\mathbb{P}(T'_{J'} = V^0, S'_{J'} = s^* | T') \geq \beta_4 > 0$ . The probability of accepting the obtained skeleton  $S' = s^*$  at stage (S2) is clearly equal to  $\beta_1$ . We have shown that  $\mathbb{P}(V^0 \in \mathcal{T}, T'_{J'} = V^0, S'_{J'} = s^* | X) \geq \beta_1 \beta_4 \beta_3 = \beta > 0$  whenever  $|J(X)| \leq h$ . Moreover, if  $T'_{J'} = V^0$  and  $S'_{J'} = s^*$  then the probability distribution of  $(T'_{J'}, S'_{J'}) \equiv X'$  does not depend on  $X$ . This completes the proof.

**PROOF OF THEOREM 1.** Theorem 1 immediately follows from Propositions 1.1 and 1.2, see for example (Roberts and Rosenthal, 2004, Th. 9).

## References

- Boys, R.J., Wilkinson, D.J., Kirkwood, T.B., 2008. Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing* 18, 125–135.
- El-Hay, T., Friedman, N., Kupferman, R., 2008. Gibbs sampling in factorized continuous-time Markov processes, in: *Proceedings of the Twenty-Fourth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-08)*, AUAI Press, Corvallis, Oregon. pp. 169–178.
- Fan, Y., Shelton, C.R., 2008. Sampling for approximate inference in continuous time Bayesian networks, in: *Tenth International Symposium on Artificial Intelligence and Mathematics*.
- Fearnhead, P., Sherlock, C., 2006. An exact Gibbs sampler for the Markov-modulated Poisson process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68, 767–784. doi:[10.1111/j.1467-9868.2006.00566.x](https://doi.org/10.1111/j.1467-9868.2006.00566.x).

- Golightly, A., Henderson, D., Sherlock, C., 2015. Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Statistics and Computing* 25, 1039–1055. URL: <http://dx.doi.org/10.1007/s11222-014-9469-x>, doi:10.1007/s11222-014-9469-x.
- Golightly, A., Wilkinson, D.J., 2011. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus* doi:10.1098/rsfs.2011.0047.
- Golightly, A., Wilkinson, D.J., 2014. Bayesian inference for Markov jump processes with informative observations. *ArXiv e-prints* .
- Jensen, A., 1953. Markoff chains as an aid in the study of Markoff processes. *Scandinavian Actuarial Journal* 1953, 87–91.
- Nodelman, U., Shelton, C.R., Koller, D., 2002. Learning continuous time Bayesian networks, in: *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers Inc.. pp. 451–458.
- Rao, V., Teh, Y.W., 2012. MCMC for continuous-time discrete-state systems, in: *Advances in Neural Information Processing Systems*, pp. 701–709.
- Rao, V., Teh, Y.W., 2013. Fast MCMC sampling for Markov jump processes and extensions. *Journal of Machine Learning Research* 14, 3207–3232.
- Roberts, G.O., Rosenthal, J.S., 2004. General state space Markov chains and MCMC algorithms. *Probability Surveys* 1, 20–71.